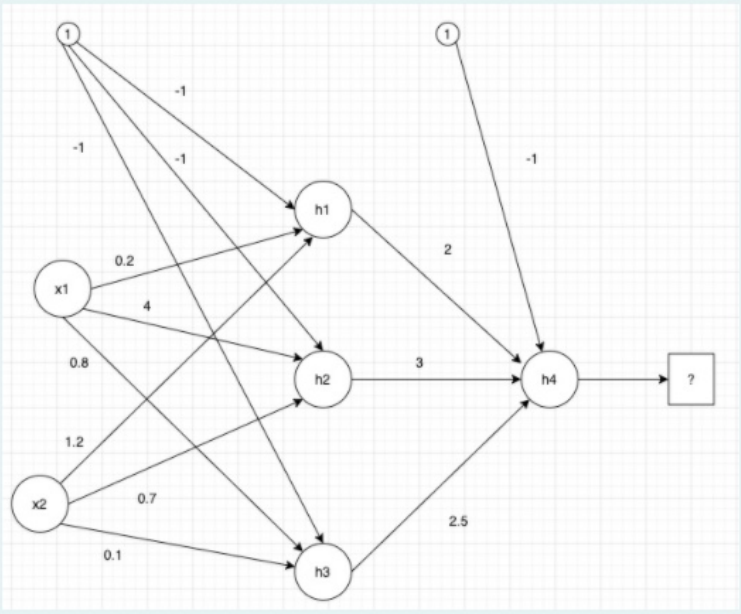
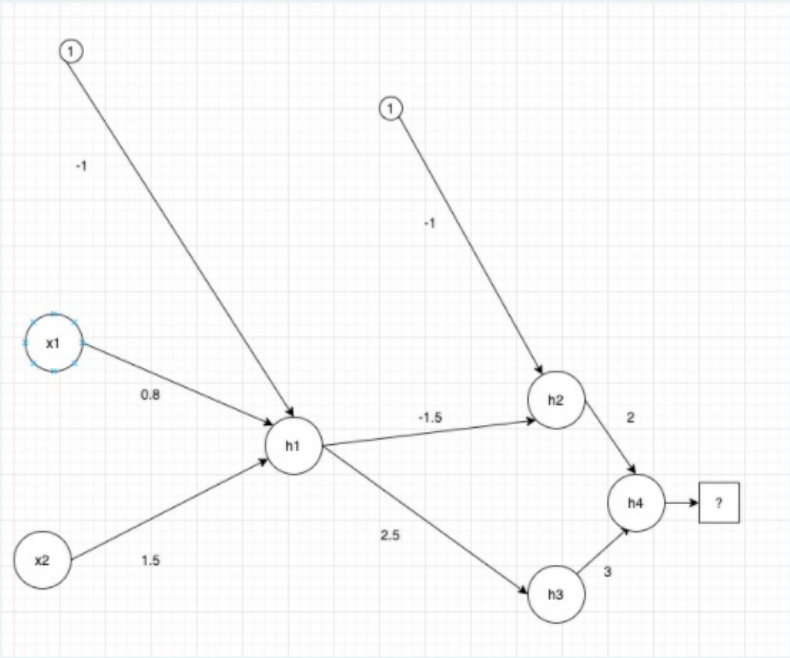
Consider the following network and assume we are using ReLU activation ($\sigma(z) = \max(0, z)$). What is the output of the fourth hidden unit h_4 when we have inputs $x_1 = 1, x_2 = 2$?



a. 15.1
b. 15.9
c. 16.4
d. 15.4
e. 15.3

Consider the following network and assume we are using ReLU activation ($\sigma(z) = \max(0, z)$). What is the output of the fourth hidden unit h_4 when $x_1 = 1, x_2 = 0.5$?



a. 4.125
b. 3.125
c. 4.5
d. 3.5
e. 5.25

4.1

You have two classes of 1 and -1 and you run AdaBoost with a logistic regression base classifier for three iterations. After doing so, you have three logistic regression models m_1, m_2, m_3 which achieve weighted errors $\epsilon_1 = 0.3, \epsilon_2 = 0.25, \epsilon_3 = 0.41$ respectively. For a new test point, the predictions of your three models are $0.4, 0.86, 0.56$ respectively. What is the prediction of your boosted model when you use a threshold of 0.6 for the base models?

- a. undefined
- b. +1
- c. -1 ✔
- d. 0 (indifferent between the two classes)

Your answer is correct.

Threshold of 0.6 means the predictions are {-1, 1, -1} for the three models respectively. The alphas (weights) of each of the three models are then {0.4236, 0.5493, 0.1820} respectively. The output of the model is then sign(-1 * 0.4236 + 0.5493 -1*0.182) = sign(-0.0563) = -1

The correct answer is: -1

A multilayer perceptron consists of an input of size 5, fully connected to a hidden layer of size 20, fully connected to an output of size 3.

Given a bias is used at each hidden node and each output node, what is the total number of parameters that need to be learned for this network?

- a. 163
- b. 165
- c. 183 ✔
- d. 185

Mila wishes to build a machine learning model to classify PET images as either containing cancerous tumours, or not. Mila performs the following steps:

- First, she splits her data into training/test sets.
- Next, Mila creates 100 subsets of her training data by resampling images from the training data, with replacement, until each dataset contains 200 images.
- Then, she trains 100 models (one for each subset, and each model is a logistic regression)
- Finally, Mila runs each of the 100 models through the test set, and uses as her final prediction for each test image the majority vote of the 100 models.

Mila's experiment is an example of the ensemble method known as:

- a. Bootstrapping
- b. Boosting
- c. Adaptive Boosting
- d. Bagging ✔

Mila wishes to build a machine learning model to classify PET images as either containing cancerous tumours, or not. Mila chooses to use decision trees as her base model, and performs the following steps:

- First, she splits her data into training/test sets.
- Next, Mila creates 100 subsets of her training data by resampling images from the training data, with replacement, until each dataset contains 200 images.
- Then, for each subset, she chooses to only look at a random sample of 40 pixels (same pixels for images in the same subset)
- Then, she trains 100 models (one for each subset)
- Finally, Mila runs each of the 100 models through the test set, and uses as her final prediction for each test image the majority vote of the 100 models.

Mila's experiment is an example of the ensemble method known as:

- a. Bagging
- b. Bootstrapping
- c. Boosting
- d. Random Forest ✔